Source:

Having a valid and reliable source to find data for a machine learning project is essential. One of our main goals for this project is to determine the effectiveness of weather data as a predictor for flight delays. To carefully find the relationship between weather and flight delays, we needed a dataset that contained both the weather data and the flight data. However, after exploring online resources for such data, we were unable to find such a dataset. The subsequent option available to find the data was to find one source for weather data and one for flight data.

We found the Department Of Transportation (DOT) a credible source for flight data. The website allowed us to extract many important variables related to flight information such as flight date, time, airport identification number, departure time, the number of minutes the flight is delayed (difference in minutes between scheduled and actual departure), and delay cause (cause of the delay such weather delay, carrier delay, security delay, and late aircraft delay). It is obvious the delay could happen for multiple reasons but for the sake of our project we will only focus on the delays related to weather.

While there are many options available to find weather data such as weather underground, and NOAA, we found Iowa Environmental Mesonet (IEM) a reliable source for the following two reasons. First, IEM soleley focuses on airport weather data, not only in the US but also airports around the world. IEM extracts the data through Automated Surface Observing System.

Then here talk about how did you combine them…..

To be continued

Questions:

What type of joining is it ?
We matched the up by the hour?